

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 10307580 A

(43) Date of publication of application: 17 . 11 . 98

(51) Int. Cl. G10H 1/00
G10L 3/00
G10L 3/00
G10L 3/00

(21) Application number: 09115408

(22) Date of filing: 06 . 05 . 97

(71) Applicant: NIPPON TELEGR & TELEPH
CORP <NTT>

(72) Inventor: MINAMI KENICHI
AKUTSU AKITO
TONOMURA YOSHINOBU

(54) MUSIC SEARCHING METHOD AND DEVICE

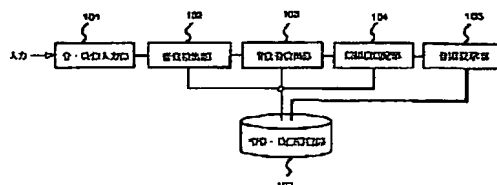
each other by a high speed bus.

(57) Abstract:

COPYRIGHT: (C)1998,JPO

PROBLEM TO BE SOLVED: To efficiently search music information by detecting a section containing music from the A/D converted, music information accumulating start/finish points of a section containing only music, measuring similarity of accumulated music, and displaying music having high similarity by a generating frequency.

SOLUTION: A music searching device is composed of a music detecting part 102, a background noise detecting part 103, a similarity measuring part 104, a music display part 105 and a music-image accumulating part 106. The music detecting part 102 detects music from inputted music information. The background noise detecting part 103 detects a sound except for music. The similarity measuring part 104 measures similarity of a section in which only music exists. The music display part 105 displays music having high similarity in order of similarity, and reproduces designated music. The music-image accumulating part 106 accumulates start/finish points of a section containing only detected music. In order to read out music at a high speed at reproducing time, the music-image accumulating part 106 and the music display part 105 are connected to



BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-307580

(43) 公開日 平成10年(1998)11月17日

(51) IntCl.⁶

識別記号

F I

G 1 0 H 1/00

G 1 0 H 1/00

Z

G 1 0 L 3/00

G 1 0 L 3/00

A

5 3 1

5 3 1 N

5 5 1

5 5 1 G

審査請求 未請求 請求項の数14 O L (全 8 頁)

(21) 出願番号

特願平9-115408

(22) 出願日

平成9年(1997)5月6日

(71) 出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72) 発明者 南 益一

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72) 発明者 阿久津 明人

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72) 発明者 外村 佳伸

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

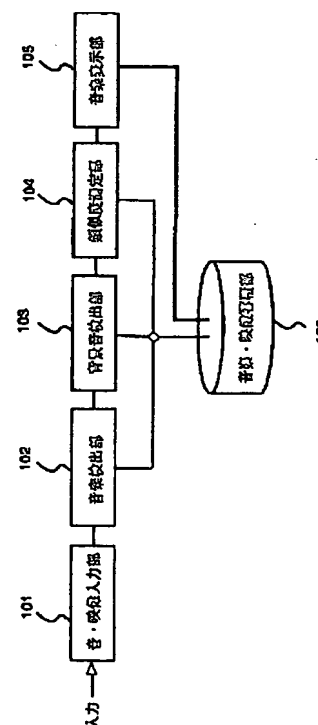
(74) 代理人 弁理士 小笠原 吉義 (外1名)

(54) 【発明の名称】 音楽検索方法および装置

(57) 【要約】

【課題】本発明は、音情報の音楽のみの区間について拍子を測定し、大局的な類似性で測定対象を絞り込んだ後、周波数スペクトルの相関を用いて効率的に音楽の整合および検索を可能にすることを目的としている。

【解決手段】デジタルの形にした入力から、音楽の含まれる区間を抽出し、その上で背景音を除いた音楽のみの区間を見出し、当該音楽の区間を用いて拍子を測定してこれを利用し、類似度を測定して音楽を検索するようにする。



BEST AVAILABLE COPY

し、前記アイコンを指定することにより対応する音楽が再生可能なことを特徴とする請求項8、9、10、11、12または13記載の音楽検索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音楽情報検索の効率化を図った音楽検索方法および装置に関する。

【0002】

【従来の技術】音楽情報を検索する際には、曲名、作詩・作曲者名、歌手名といったキーワードを用いる方法があるが、キーワードの付与を自動的に行うことは難しく、大量の情報を扱う場合にはデータベースの構築に多大な労力を要する。また、曲名等が不明な場合の検索が不可能である。

【0003】このような問題を解決するために、音楽の旋律情報を手がかりとして検索方法が報告されており、パーソナルコンピュータのキーボードを鍵盤楽器に見立て、旋律を入力するもの（石塚ら：“パーソナルコンピュータによる音楽情報検索システムの開発1．旋律検索”、図書館情報大学研究報告、Vol. 6, No. 1, pp. 95-105, 1987）や、曲を口ずさみ、そのピッチを旋律として入力するもの（“蔭山ら：ハミング歌唱を手掛かりとするメロディ検索、電子情報通信学会論文誌、D-II, Vol. J77, No. 8, pp. 1543-1551, 1994）がある。入力と音楽データベースとの整合には、旋律情報の動的プログラミング（DP）を行う方法を用いている。

【0004】

【発明の解決しようとする課題】日常生活において音楽の情報を得る機会、テレビ放送による場合が少なくない。繰り返し流れる音楽を聞いているうちにその曲に対する印象を持つようになり、視聴の欲求を喚起する場合がある。しかし、うろ覚えのメロディ（旋律）程度の情報しか持たない場合には旋律の入力が困難であり、所望の音楽を特定することは難しい。更にテレビ放送では、音楽に会話やナレーションなどの音声を重ねる場合が多いため、旋律を自動的かつ正確に検出できる対象は限定されてしまう。また、旋律を用いる方法では、同じ曲であっても旋律の時間的な変化が異なる場合に対処するために、時間方向のずれを許容するようなDPが必要であり、計算コストを増加させる原因となっている。

【0005】本発明の目的は、音情報の音楽のみの区間について拍子を測定し、大局的な類似性で測定対象を絞り込んだ後、周波数スペクトルの相関を用いて効率的に音楽の整合および検索が可能にすることにある。

【0006】

【課題を解決するための手段】上記目的を達成するために本発明の請求項1記載の音楽検索方法は、音情報がアナログの場合あるいは音情報を伴う映像情報がアナログの場合にはA/D変換し、音情報あるいは音情報を伴う

映像情報がデジタルの場合にはそのまま入力する音・映像入力段階と、入力された信号から音楽が含まれる区間を検出する音楽検出段階と、該音情報から音楽以外の音が含まれる区間を検出する背景音検出段階と、音楽が含まれる区間の音情報、および音楽のみが含まれる区間の開始/終了点、および入力映像の場合には音楽が含まれる区間の映像ならびに音情報を蓄積する音楽・映像蓄積段階と、蓄積された音楽の類似度を測定する類似度測定段階と、類似度の高い音楽を発生頻度順に表示する音楽表示段階とを有することを特徴とする。

【0007】また、本発明の請求項2記載の音楽検索方法は、音楽検出段階において、該音情報を周波数解析した際のスペクトルのピークが周波数方向に時間的に安定していることを利用して音楽を検出することを特徴とする。

【0008】また、本発明の請求項3記載の音楽検索方法は、背景音検出段階において、くし形フィルタを用いて音声の有声音に含まれる調波構造を検出し、背景音の1つである音声を検出することを特徴とする。

【0009】また、本発明の請求項4記載の音楽検索方法は、背景音検出段階において、入力された音情報から得られた周波数ケプストラム係数と符号帳とのユークリッド距離が近いものを背景音として検出することを特徴とする。

【0010】また、本発明の請求項5記載の音楽検索方法は、類似度測定段階において、蓄積された音情報の音楽のみが含まれる区間における周波数スペクトルパワーの立ち上がり周期から音楽の拍子を測定し、同等の拍子を持つものだけを類似度の測定対象とすることを特徴とする。

【0011】また、本発明の請求項6記載の音楽検索方法は、類似度測定段階において、蓄積された音情報の音楽のみが含まれる区間における周波数スペクトルの相互相関から、あるいは前記音情報の音楽のみが含まれる区間の自己相関から得られたピッチの変化パターンの相互相関から蓄積された音情報同士の類似度を測定することを特徴とする。

【0012】また、本発明の請求項7記載の音楽検索方法は、音楽表示段階において、類似度の高い音楽を順にアイコンで表示するようにし、かつ入力映像の場合には音楽開始点に対応する画像をアイコンとして表示し、前記アイコンを指定することにより対応する音楽を再生することを特徴とする。

【0013】さらに、情報目的を達成するために本発明の請求項8記載の音楽検索装置は、音情報がアナログの場合あるいは音情報を伴う映像情報がアナログの場合にはA/D変換し、音情報あるいは音情報を伴う映像情報がデジタルの場合にはそのまま入力する音・映像入力部と、入力された信号から音楽が含まれる区間を検出する音楽検出部と、音情報から音楽以外の音が含まれる区

間を検出する背景音検出部と、音楽が含まれる区間の音情報、および音楽のみが含まれる区間の開始/終了点、および入力映像の場合には音楽が含まれる区間の映像ならびに音情報を蓄積する音楽・映像蓄積部と、蓄積された音楽の類似度を測定する類似度測定部と、類似度の高い音楽を発生頻度順に表示する音楽表示部とを具備することを特徴とする。

【0014】また、本発明の請求項9記載の音楽検索装置は、音楽検出部において、音情報を周波数解析した際のスペクトルのピークが周波数方向に時間的に安定していることを利用して音楽を検出することを特徴とする。

【0015】また、本発明の請求項10記載の音楽検索装置は、背景音検出部において、くし形フィルタを用いて音声の有声音に含まれる調波構造を検出し、背景音の1つである音声を検出することを特徴とする。

【0016】また、本発明の請求項11記載の音楽検索装置は、背景音検出部において、入力された音情報から得られた周波数ケプストラム係数と符号帳とのユークリッド距離が近いものを背景音として検出することを特徴とする。

【0017】また、本発明の請求項12記載の音楽検索装置は、類似度測定部において、蓄積された音情報の音楽のみが含まれる区間における周波数スペクトルパワーの立ち上がり周期から音楽の拍子を測定し、同等の拍子を持つものだけを類似度の測定対象とすることを特徴とする。

【0018】また、本発明の請求項13記載の音楽検索装置は、類似度測定部において、蓄積された音情報の音楽のみが含まれる区間における周波数スペクトルの相互相関から、あるいは前記音情報の音楽のみが含まれる区間の自己相関から得られたピッチの変化パターンの相互相関から蓄積された音情報同士の類似度を測定することを特徴とする。

【0019】また、本発明の請求項14記載の音楽検索装置は、音楽表示部において、類似度の高い音楽を順にアイコンで表示するようにし、かつ入力映像の場合には音楽開始点に対応する画像をアイコンとして表示し、前記アイコンを指定することにより対応する音楽が再生可能なことを特徴とする。

【0020】請求項1および8記載の音楽検索方法および装置は、音情報がアナログの場合あるいは音情報を伴う映像情報がアナログの場合にはA/D変換し、音情報あるいは音情報を伴う映像情報がデジタルの場合にはそのまま入力し、入力された音情報から音楽が含まれる区間を検出し、音情報から音楽以外の音が含まれる区間を検出し、音楽が含まれる区間の音情報、および音楽のみが含まれる区間の開始/終了点、および入力映像の場合には音楽が含まれる区間の映像ならびに音情報を蓄積し、蓄積された音楽の類似度を測定し、類似度の高い音楽を発生頻度順に表示することにより、繰り返し入力

される音楽を発生頻度に基づいて検索することが可能となる。

【0021】請求項2および9記載の音楽検索方法および装置は、音情報を周波数解析した際のスペクトルのピークが周波数方向に時間的に安定していることを利用することにより音楽を検出することが可能となる。

【0022】請求項3および10記載の音楽検索方法および装置は、くし形フィルタを用いて音声の有声音に含まれる調波構造を検出することにより背景音の1つである音声を検出することが可能となる。

【0023】請求項4および11記載の音楽検索方法および装置は、入力された音情報から得られた周波数ケプストラム係数と符号帳とのユークリッド距離を算出することにより距離が近いものを背景音として検出することが可能となる。

【0024】請求項5および12記載の音楽検索方法および装置は、蓄積された音情報の音楽のみが含まれる区間における周波数スペクトルパワーの立ち上がり周期から音楽の拍子を測定し、同等の拍子を持つものだけを類似度の測定対象とすることにより検索対象の絞り込みが可能となる。

【0025】請求項6および13記載の音楽検索方法および装置は、蓄積された音情報の音楽のみが含まれる区間における周波数スペクトルの相互相関から、あるいは前記音情報の音楽のみが含まれる区間の自己相関から得られたピッチの変化パターンの相互相関を用いることにより蓄積された音情報同士の類似度を測定することが可能となる。

【0026】請求項7および14記載の音楽検索方法および装置は、類似度の高い音楽を順にアイコンで表示するようにし、かつ入力映像の場合には音楽開始点に対応する画像をアイコンとして表示することにより視覚的に音楽を選択・再生することが可能となる。

【0027】

【発明の実施の形態】次に、本発明の実施の形態について図面を参照して説明する。図1は本発明の一実施形態の音楽検索装置の概略構成を示すブロック図である。本実施形態の音楽検索装置は、音あるいは映像（音を含む）を入力する音・映像入力部101と、入力された音情報から音楽を検出する音楽検出部102と、音楽以外の音を検出する背景音検出部103と、音楽のみが存在する区間の類似度を測定する類似度測定部104と、類似度の高い音楽を頻度順に表示し、指定された音楽を再生する音楽表示部105と、検出された音楽、映像、音楽のみが含まれる区間の開始/終了点を蓄積する音楽・映像蓄積部106とから構成されている。また、再生時に音楽あるいは映像を高速に読み出すため、音楽・映像蓄積部106と音楽表示部105とは、高速なバスで接続されている。

【0028】図2は音楽と背景音との物理的な性質の差

異を説明する図である。音楽のような周期性のある信号のスペクトラムは、図2(a)に示す如く多くの特定の周波数において急峻に立ち上がるピークをもつものとなる。また逆に、背景音の場合には、そのスペクトラムは、図2(b)に示す如く所定の周波数帯域内で拡がりをもちかつ特定の周波数において小さいピークをもつものとなる。

【0029】図2(a)に示すようなスペクトラムをもつ信号と図2(b)に示すようなスペクトラムをもつ信号との夫々のケプストラム係数をとると、図2(a)の場合のスペクトル包絡や図2(b)の場合のスペクトル包絡を近似的に表現する値となることから、両者の特徴の違いを定量的に評価することができる。

【0030】なおケプストラム係数は、処理対象の信号についてフーリエ変換をとり、その対数値を計算し、その上で逆フーリエ変換をとる処理を行うことによって得られる。なお上記フーリエ変換の代わりに線形予測法を用いることもできる。

【0031】図3は本発明の一実施形態の処理の流れを示すフローチャートである。本発明をソフトウェアで実現した場合でも同様の処理の流れとなる。まず、入力された音情報を処理201においてフーリエ変換し、パワースペクトルを算出する。次にパワースペクトルの周波数方向の微分値を取るによりパワースペクトルのピークの位置を算出し、数秒間の平均持続時間PTを処理202において算出する。

【0032】図4は音楽におけるスペクトルピークの持続性を説明する図である。音楽のサウンドスペクトログラムにおいて夫々の周波数毎の音圧の高い所即ち夫々の周波数でのピークが時間的にどのように持続するかを模式的に表すと、図4に示す如きものとなる。図4において、横軸は時間 t であり、縦軸は周波数 f であり、個々の周波数毎に音圧の高い所(ピーク)が持続している状況が図示の水平方向の実線で表現されている。

【0033】この実線の長さ(時間的な長さ)が大である程、当該ピークの持続時間が大きいことを表している。したがって、図4に示す如く所定の時間間隔の窓を考えて、当該窓の中において、図示の実線の長さの平均値(平均持続時間PT)をとると、当該平均値の大きい窓の所では音楽が含まれているものと判断される。

【0034】平均持続時間PTが閾値TH1を越えた場合に入力された音情報に音楽が含まれると判断し、音情報の蓄積を開始し、また入力に映像がある場合にはその映像をも蓄積を開始する(処理203)。平均持続時間PTが閾値以下の場合には、処理211において蓄積を終了する。

【0035】次にくし形フィルタの出力CRを処理204において算出する。くし形フィルタは、周波数方向に等間隔に配置された複数の帯域フィルタであるので、声に含まれる調波構造を検出することで人の声を検出する

ことが可能である。

【0036】調波構造は音楽においても観測されるため、くし形フィルタは、音楽のスペクトル成分を除去した後に適用される。処理202において算出されたパワースペクトルのピークのうち、周波数方向の時間的な変動が少ないものの値を強制的に零とし、その周囲のスペクトルの値を用いて零にされたスペクトルの値を補完するようにする。音声のスペクトルは時間的な変動が激しいので保存されることになる。

【0037】複数の帯域フィルタの夫々の出力の和である出力CRが閾値TH2以上の時、音情報に声が存在すると判断され、背景音の存在する期間が開始したとして処理210において背景音開始点を蓄積する。CRが閾値TH2以下の場合には、一般には音楽の存在する期間に当たるものであるがなお背景音が含まれることがある。このために雑踏等様々な背景音を学習データとして用いて予め生成されたコードブック(符号帳)と入力された音情報のスペクトルの特徴量とのユークリッド距離EDを処理205において算出する。特徴量には、周波数ケプストラム等を用いる。

【0038】EDが閾値TH3以上の場合には、背景音の存在する期間とみて背景音開始点を処理210において蓄積し、閾値以下の場合には、背景音終了点を処理206において蓄積する。

【0039】このようにして入力が終わった後は、いわば音楽のみの期間について当該音楽の期間の抽出が終了する。次いで、蓄積された音楽の音楽のみの部分を用いて拍子の測定を処理207において行う。拍子の測定に当たっては、スペクトルの急激な立ち上がりを検出し自己相関係数を求めることによって行う。

【0040】次いで各々の音楽を数段階の拍子レベルに分類し、同じレベルに属するものを対象にして類似度を測定する(処理208)。類似度の測定には、一方のスペクトルと、もう一方のスペクトルの相互相関係数を求めることにより行う。また、音楽のピッチを検出してその変化のパターンの相関係数を求めることにより行うことも可能である。相関係数の算出過程においては、自乗誤差が大きくなった場合に計算を打ち切り、処理時間を低減するSSDA法等を併用する。相関が高い音楽区間の数を頻度とし、頻度の高い順に音楽区間に対応する映像を静止画像のアイコンとして処理209において表示する図5は、インタフェース上に表示されたアイコンの様子を示している。アイコン301上には、音楽区間の先頭に対応する画像が表示されるようにされており、スクロールバー302、303を用いて枠外のアイコンを表示させる。マウス等のポインティングデバイスを用いてアイコンを指定することにより、対応する音楽あるいは映像が再生される。入力が映像ではなく音情報のみの場合には、疑似的なアイコンを用いて表示することが可能である。

【0041】

【発明の効果】

(1) 請求項1および8の発明は、入力情報を必要に応じてA/D変換し、入力された音情報から音楽が含まれる区間を検出し、音情報から音楽以外の音が含まれる区間を検出し、音楽が含まれる区間の音情報ならびに映像情報、および音楽のみが含まれる区間の開始/終了点を蓄積し、蓄積された音楽の類似度を測定し、類似度の高い音楽を発生頻度順に表示することにより、繰り返し入力される音楽を発生頻度に基づいて検索することができる。

(2) 請求項2および9の記載の発明は、音情報を周波数解析した際のスペクトルのピークが周波数方向に時間的に安定していることを利用することにより音楽を検出することができる。

(3) 請求項3および10の発明は、くし形フィルタを用いて音声の有声音に含まれる調波構造を検出することにより背景音の1つである音声を検出することができる。

(4) 請求項4および11の記載の発明は、入力された音情報から得られた周波数ケプストラム係数と符号帳とのユークリッド距離を算出することにより距離に近いものを背景音として検出することができる。

(5) 請求項5および12の記載の発明は、蓄積された音情報の音楽のみが含まれる区間における周波数スペクトルパワーの立ち上がり周期から音楽の拍子を測定し、同等の拍子を持つものだけを類似度の測定対象とすることにより検索対象の絞り込みができる。

(6) 請求項6および13の記載の発明は、蓄積された音情報の音楽のみが含まれる区間における周波数スペクトルの相互相関から、あるいは前記音情報の音楽のみが含まれる区間の自己相関から得られたピッチの変化パターンの相互相関を用いることにより蓄積された音情報同士の類似度を測定することができる。

(7) 請求項7および14の記載の発明は、類似度の高

い音楽を順にアイコンで表示するようにし、かつ入力映像の場合には音楽開始点に対応する画像をアイコンとして表示することにより視覚的に音楽を選択・再生することができる。

【図面の簡単な説明】

【図1】本発明の一実施形態の音楽検索装置の概略構成を示すブロック図である。

【図2】音楽と背景音との物理的な性質の差異を説明する図である。

【図3】本発明の一実施形態の処理の流れを示すフローチャートである。

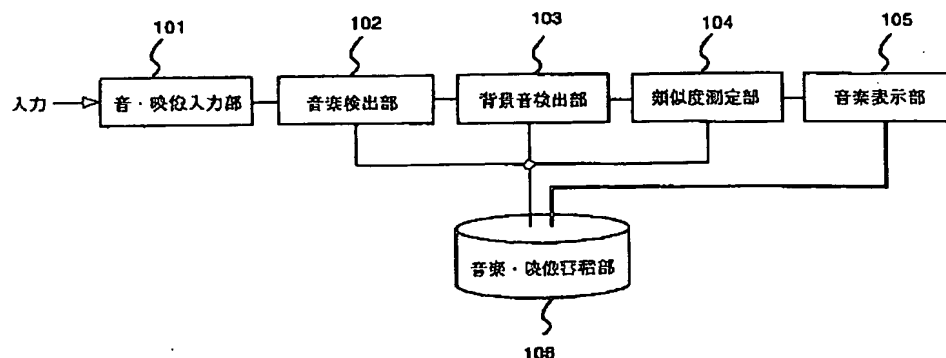
【図4】音楽におけるスペクトルピークの持続性を説明する図である。

【図5】本発明の一実施形態の音楽検索装置の音楽表示部の外観を示す図である。

【符号の説明】

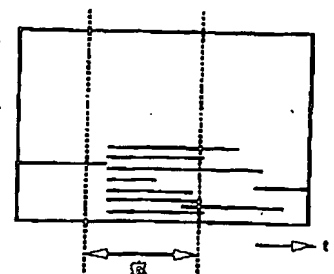
- 101 音・映像入力部
- 102 音楽検出部
- 103 背景音検出部
- 104 類似度測定部
- 105 音楽表示部
- 106 音楽・映像蓄積部
- 201 FFT処理
- 202 ピーク持続時間算出処理
- 203 音・映像蓄積開始処理
- 204 くし形フィルタ適用処理
- 205 ユークリッド距離算出処理
- 206 背景音終了点蓄積処理
- 207 拍子測定処理
- 208 類似度測定処理
- 209 音楽表示処理
- 210 背景音開始点蓄積処理
- 211 音・映像蓄積終了処理
- 301 アイコン
- 302, 303 スクロールバー

【図1】



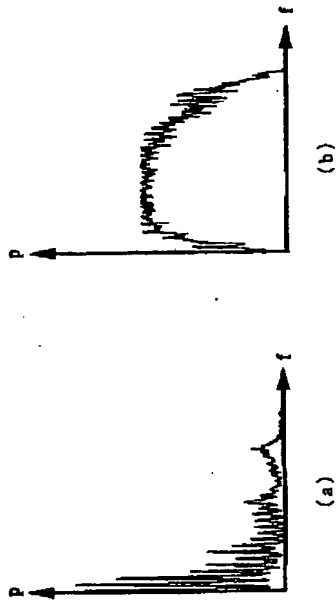
【図4】

音楽におけるスペクトルピークの持続性

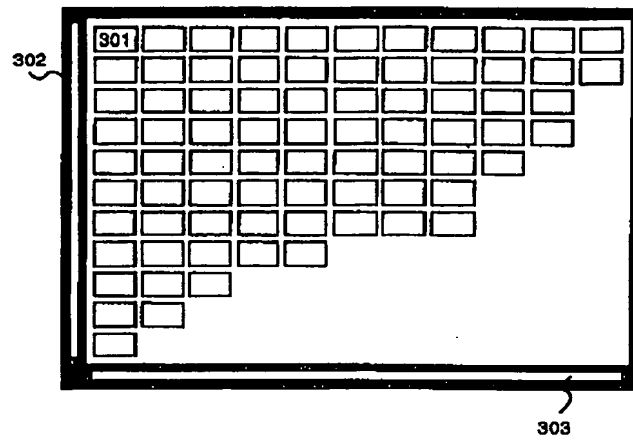


【図2】

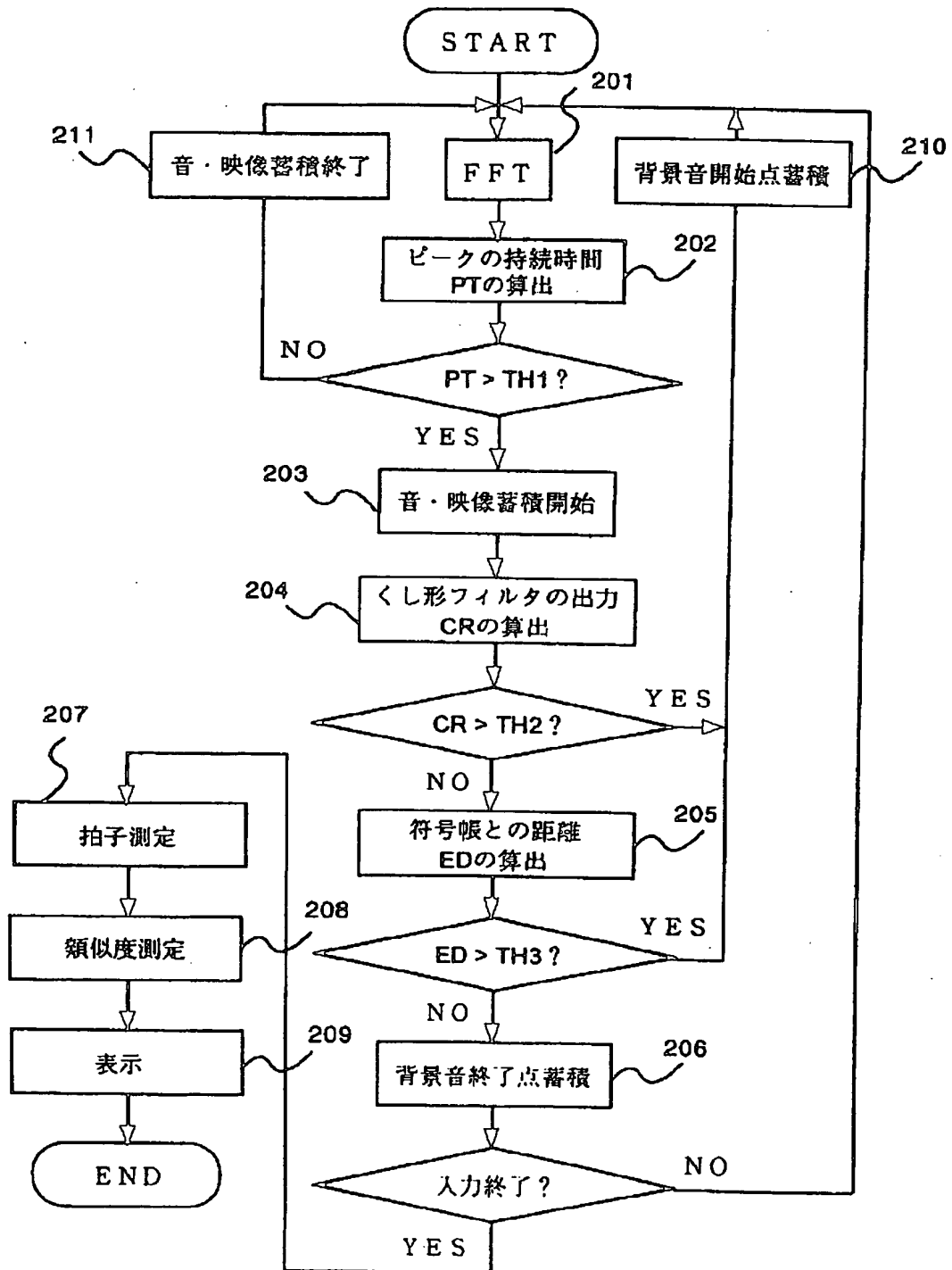
音楽と背景音



【図5】



【図3】



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.